University of Toronto Mississauga

STA258H5 - Winter 2018
R Assignment for Term Test # 2

Instructors: Dr. Alison Weir and Dr. Al Nosedal

Problem 1. Refer to Kimberly-Clark Corporation's survey of 250 people who kept a count of their use of Kleenex® tissues in diaries. We want to test the claim made by marketing experts that $\mu = 60$ is the average number of tissues used by people with colds against our belief that the population mean is smaller than 60 tissues. We will select $\alpha = 0.05$ as the level of significance for the test.
The survey results for the 250 sampled Kleenex® users are in the *tissues.txt* data file. Data are available at

```
tissues_url = "http://www.math.unm.edu/~alvaro/tissues.txt"
```

Please do the following:

a) Assign the values from the variable *NUMUSED* to an object (vector) that bears your family name. Example:

```
tissues_url = "http://www.math.unm.edu/~alvaro/tissues.txt"
tissues_data = read.table(tissues_url,header=TRUE);
blackwell = tissues_data$NUMUSED;
```

b) Using R, conduct a test at 5% significance level.

c) State your conclusion in the context of this problem.

Put your answers to parts a) - c) in a word document that fits on **one sheet of paper**. At the top of the page, put your name and student number. Title the page *Tissues Data (Mean)*. Copy-and-paste your commands and the responses from the session window to the word document.

Problem 2. In the previous problem, we investigated Kimberly-Clark Corporation's assertion that the company should put 60 tissues in a cold-care box of Kleenex® tissues. Another approach to the problem is to consider the proportion of Kleenex® users who use fewer than 60 tissues when they have a cold. Now the population parameter of interest is p, the proportion of all Kleenex® users who use fewer than 60 tissues when they have a cold.
Kimberly-Clark Corporation's belief that the company should put 60 tissues in a cold-care box will be supported if the *median* number of tissues used is 60. Now, if the true median is, in fact, 60, then half of the Kleenex® users will use less than 60 tissues (i.e., $p = 0.5$). Is there evidence to indicate that the population proportion differs from 0.5? We will select $\alpha = 0.05$ as the level of significance for the test.

Recall that the survey results for the 250 sampled Kleenex® users are stored in the *tissues.txt* data file. In addition to the number of tissues used by each person, the file contains a qualitative variable - called USED60 - representing whether the person used fewer or more than 60 tissues. (The values of USED60 in the data set are BELOW=0 or ABOVE=1.)

Please do the following:

a) Assign values from variable *USED60* to an object (vector) that bears your family name.

b) Using R, conduct a test at 5% significance level.

c) State your conclusion in the context of this problem.

Put your answers to parts a) - c) in a word document that fits on **one sheet of paper**. At the top of the page, put your name and student number. Title the page *Tissues Data (Proportion)*.
Copy-and-paste your commands and the responses from the session window to the word document.

Problem 3. Behavioral researchers have developed an index designed to measure managerial success. The index (measured on a 100-point scale) is based on the manager's length of time in the organization and his or her level within the firm; the higher the index, the more successful the manager. Suppose a researcher wants to compare the average success index for two groups of managers at a large manufacturing plant. Managers in group 1 engage in a high volume of interactions with people outside the manager's work unit. (Such interactions include phone and face-to-face meetings with customers and suppliers, outside meetings, and public relations work.) Managers in group 2 rarely interact with people outside their work unit. Independent random samples of 12 and 15 managers are selected from groups 1 and 2, respectively, and the success index of each is recorded. The results of the study are available at:

```
index_url = "http://www.math.unm.edu/~alvaro/mansuccess.txt"
```

Please do the following:

a) Use R and the data in *mansuccess.txt* to estimate the true mean difference between the success indexes of managers in the two groups. Use a 95% confidence interval.

b) Interpret the interval from part a).

c) What assumptions must be made in order that the estimate be valid? Are they reasonably satisfied?

Put your answers to parts a) and c) in a word document that fits on **two sheets of paper**. At the top of the pages, put your name and student number. Title the pages *Index Data*. Copy-and-paste your commands and the responses from the session window to the word document.
Please, resize plots so that all of them fit on one side of an 8.5 x 11 sheet of paper. Title your plots with your family name.

Problem 4. To determine the effect of advertising in the Yellow Pages, Bell Telephone took a random sample of 40 retail stores that did not advertise in the Yellow Pages last year but did so this year. The annual sales (in thousands of dollars) for each store in both years were recorded. Data set is available at:

```
sales_url = "http://www.math.unm.edu/~alvaro/sales.txt"
```

Please do the following:

a) Assign the values from variable *This.Year* to an object (vector) that bears your family name and the word *current*. Example:

```
blackwell.current = sales_data$This.Year;
```

Then, assign values from variable *Last.Year* to another object (vector) that bears your family name and the word *old*. Example:

```
blackwell.old = sales_data$Last.Year;
```

b) Use R and the data in *sales.txt* to estimate with 90% confidence the improvement in sales between the two years.

c) Check to ensure that the required condition(s) of the technique used in part b) is satisfied.

d) Would it be advantageous to perform this experiment with independent samples? Explain why or why not.

Put your answers to parts a) and d) in a word document that fits on **one sheet of paper**. At the top of the page, put your name and student number. Title the page *Sales Data*. Copy-and-paste your commands and the responses from the session window to the word document.
Please, resize plots so that all of them fit on one side of an 8.5 x 11 sheet of paper. Title your plots with your family name. Copy-and-paste your commands.

Problem 5. A consumer products company is formulating a new shampoo and is interested in foam height (in millimeters). Foam height is approximately Normally distributed and has a standard deviation of 20 millimeters. The company wishes to test $H_0 : \mu = 175$ millimeters versus $H_a : \mu > 175$ millimeters, using the results of $n = 10$ samples.

Please do the following:

a) Using R, find the type I error probability α if the critical region is $\bar{x} > 185$.

b) What is the probability of type II error if the true mean foam height is 196 millimeters? **Use R** to answer this question, too.

c) If we plot the probability of "accepting" (failing to reject) $H_0 : \mu = \mu_0$ versus various values of μ and connect the points with a smooth curve, we obtain the **operating characteristic curve** (or the OC curve) of the test procedure. These curves are used extensively in industrial applications of hypothesis testing to display sensitivity and relative performance of the test. **Using R**, construct an OC curve for the test from part a). Use values of the true mean μ of $175, 178, 181, 184, 187, 190, 193, 196, 199$, and 202. Title your plot with your family name and (OC curve), for instance: "Fisher (OC curve)".

d) Using R, make a plot of the power function of the test from part a). Use values of the true mean μ of $175, 178, 181, 184, 187, 190, 193, 196, 199$, and 202. Title your plot with your family name and (Power curve), for instance: "Fisher (Power curve)".

Put your answers to parts a) and b) in a word document that fits on **one sheet of paper**. At the top of the page, put your name and student number. Title the page *Shampoo Data*. Copy-and-paste your commands and the responses from the session window to the word document.
Put your answers to parts c) and d) in a word document that fits on one side of an 8.5 x 11 sheet of paper. That is, resize plots so that both of them fit on one side of an 8.5 x 11 sheet of paper.
Copy-and-paste your commands.

Problem 6. Some traffic experts believe that the major cause of highway collisions is the differing speeds of cars. That is, when some cars are driven slowly while others are driven at speeds well in excess of the speed limit, cars tend to congregate in bunches, increasing the probability of accidents. Thus, the greater the variation in speeds, the greater will be the number of collisions that occur. Suppose that one expert believes that when the variance exceeds 18 mph^2, the number of accidents will be unacceptably high. A random sample of the speeds of 245 cars on a highway with one of the highest accident rates in the country is taken. Can we conclude at the 10% significance level that the variance in speeds exceeds 18 mph^2?

Data set is available at:

```
speeds_url = "http://www.math.unm.edu/~alvaro/speeds.txt"
```

Please do the following:

a) Assign values from variable *Speeds* to an object (vector) that bears your family name.

b) Using R, conduct a test at 10% significance level.

c) State your conclusion in the context of this problem.

d) Check to ensure that the required condition(s) of the technique used is satisfied.

Put your answers to parts a) - c) in a word document that fits on **one sheet of paper**. At the top of the page, put your name and student number. Title the page *Speeds Data*. Copy-and-paste your commands and the responses from the session window to the word document.
Put your answers to part d) in a word document that fits on **one sheet of paper**. Resize plots so that all of them fit on one side of an 8.5 x 11 sheet of paper. Title your plots with your family name. Copy-and-paste your commands.

Problem 7. A new highway has just been completed and the government must decide on speed limits. There are several possible choices. However, on advice from police who monitor traffic the objective was to reduce the variation in speeds, which it is thought to contribute to the number of collisions. It has been acknowledged that speed contributes to the severity of collisions. It is decided to conduct an experiment to acquire more information. Signs are posted for 1 week indicating that the speed limit is 70 mph. A random sample of cars' speeds is measured. During the second week, signs are posted indicating that the maximum speed is 70 mph and that the minimum speed is 60 mph. Once again a random sample of speeds is measured. Can we infer that limiting the minimum and maximum speeds reduces the variation in speeds? Use $\alpha = 0.05$.

Data set is available at:

```
new_highway_url = "http://www.math.unm.edu/~alvaro/new-highway.txt"
```

Please do the following:

a) Assign values from variable *Week1* to an object (vector) that bears your family name and the word *week1*.
Then, assign values from variable *Week2* to another object (vector) that bears your family name and the word *week2*. Example:

b) Using R, conduct a test at 5% significance level.

c) State your conclusion in the context of this problem.

d) Check to ensure that the required condition(s) of the technique used is satisfied.

Put your answers to parts a) - c) in a word document that fits on **one sheet of paper**. At the top of the page, put your name and student number. Title the page *New Highway Data*. Copy-and-paste your commands and the responses from the session window to the word document.
Put your answers to part d) in a word document that fits on **one sheet of paper**. Resize plots so that all of them fit on two sides of an 8.5 x 11 sheet of paper. Title your plots with your family name. Copy-and-paste your commands.

You should have 10 sheets (20 pages) of R output. Bring the 10 sheets (20 pages) to your second test.